Statistical Inference for Scientific Instruments: Event Analysis for the Gamma-ray Large Area Space Telescope. Grant Number NNG05GC80G – Annual Report, Year 1

Robin D. Morris, USRA-RIACS
Johann Cohen-Tanugi, SLAC

Abstract

We describe progress in developing and implementing advanced statistical methods for event analysis for the Large Area Telescope (LAT) on the Gamma-ray Large Area Space Telescope (GLAST) satellite. We describe development of the basic statistical tool – the determination of the probabilities of the elementary events within the detector; the testing of the methodology on a simplified simulation of the LAT instrument; and progress on the integration of the tools into the GLEAM (Glast Event Analysis Machine) framework. We present results on 1) a simplified simulation of the detector, used to validate the statistical methodology, and 2) showing how the methodology as implemented within GLEAM can be used to estimate the energy of muons incident on the detector, and compare the energy estimates with those from the conventional analysis. Finally, we describe future plans for how the system will be extended to allow for the analysis of more complex events, including events due to incident photons.

1 Introduction

The LAT instrument is a pair-conversion telescope. Incident photons interact with tungsten foils and are converted into electron/positron pairs. The trajectories of the charged particles are determined by observing where the charged particles traverse silicon microstrip detector layers. Figure 1 shows part of the detector from the GLEAM simulator. The track of the photon (yellow) and the charged particles (blue) can be seen, together with the microstrips that fired (cyan). The basic problem is to estimate the direction and energy of the photon.

When pair production occurs, the energy of the photon is split stochastically between the electron and the positron; the opening angle (the angle between the trajectories of the electron and positron) is also split in proportion to the energy of each of the charged particles. (See figure 2.) To determine accurately the *direction* of the photon, it is necessary therefore to determine accurately the *energy* of the electron and positron separately. This is the motivation for our study of how our methodology functions when used to estimate the energy of a single charged particle incident on the detector. See section 6.

When multiple charged particles are present, there is an inherent ambiguity in the determination of the trajectory of the electron and the trajectory of the positron. Because the silicon microstrips measure $x-$ or $y-$ position *separately*, when two charged particles traverse a layer, typically two x-strips and two y-strips will trigger. All that is know is that the two particles went through diagonally opposite corners of the rectangle defined by the strips. Which diagonal is unknown just

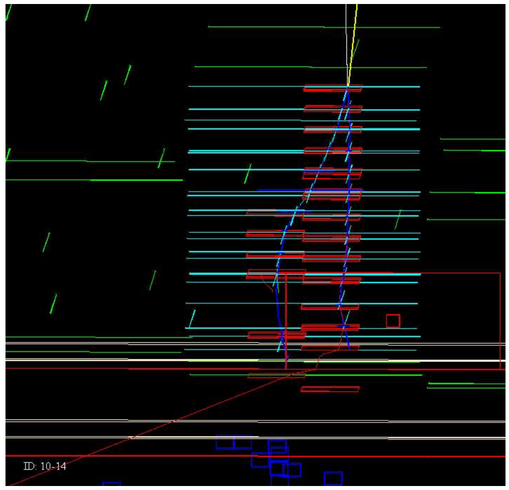

Figure 1: Simulation within GLEAM of a 100MeV photon incident on the LAT

from the response at that layer. This effect, and potential reversals of which diagonal is traversed at subsequent layers, means that there is an exponential number of possible paths for the two charged particles. The statistics of the elementary multiple scattering events can be used to determine the probability of each configuration, and to determine which configurations are sufficiently probably to be retained in the analysis. See section 3.

All steps of the analysis are dependent on the probabilities of the elementary interactions. Section 4 describes our use of the Geant4 particle physics simulation package to determine the probabilities for the outcomes when a particle is incident on one of the GLAST LAT converter foils. These probabilities are used in the later event analysis.

2 Development of the Work Plan

The initial work plan, described in the proposal, envisioned the development proceeding along the following lines. It proposed a bottom-up development plan for the first two years, during which time the basic building blocks of the approach would be developed and implemented. Only towards the end of this two year period would the blocks be assembled into a complete system that could be used to analyse events, and the system refined and tested against simulated event data and eventually actual event data.

Discussion with members of the LAT team at SLAC indicated that this timetable would not allow serious consideration of the methodology and system within the larger GLAST mission context. For the system being developed here to have a chance of being adopted by the mission for even some of the event analysis when the observatory is launched, it would be necessary to demonstrate successful analysis of events at a much earlier date, even if the type of events that could be analysed was restricted in generality.

Due to this feedback, we have switched from a purely bottom-up development approach to a mixed bottom-up/top-down approach. Instead of developing all the building blocks, we have begun by defining the required interfaces to the building blocks, and developing only a subset. This has allowed a complete analysis system to be completed within the GLEAM framework, but one that

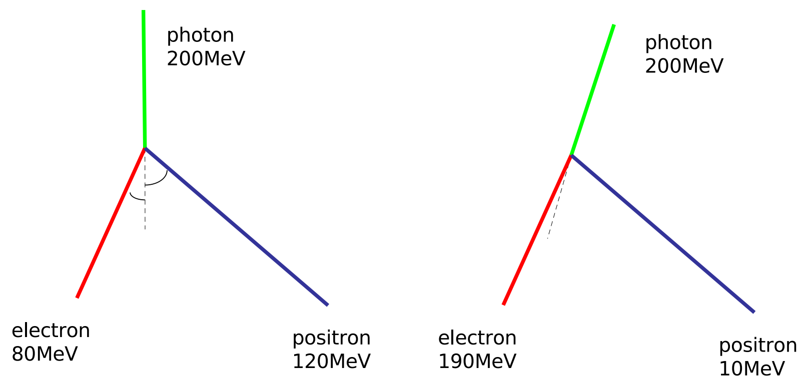

Figure 2: The determination of the direction of the photon requires the accurate determination of the energy of the electron and the energy of the positron.

currently can only analyse the interaction of muons with the detector (this is the simplest type of event in terms of the physics). This framework will now be extended to other types of charged particle (electron and positron), where the production of secondary particles in the detector is much more significant than for muons, and finally will be extended to the case of photons (gamma-rays) incident on the detector.

3 Development of a Simplified LAT Simulator

Figure 3 shows a gamma-ray (green line) incident on a series of tungsten foils (cyan squares). The foils have the same thickness and spacing as the foils in the LAT. The red and blue lines are the trajectories of the electron and positron. This simulation was performed using the Geant4 toolkit (Geant4 is also used within GLEAM as the physics simulation engine). In the proposal description, we showed results on an extremely simplified detector, using Sequential Importance Sampling (SIS) to determine the probability distribution over direction and energy by incorporating one layer at a time into the analysis, and using Markov chain Monte Carlo (MCMC) to explore the details of the distribution. It was found that on the high-fidelity simulation of the conversion foils of the detector shown in figure 3 that the details of the geometry and elementary probabilities made this approach computationally inefficient. Instead, the explicit enumeration of the possible trajectories was performed, with MCMC used to determine the distribution within each possible trajectory configuration. The probability of each configuration was estimated from the MCMC samples. Note that this can be done sequentially, after each layer is incorporated, and so configurations with negligible probability can be pruned quickly, countering the exponential growth in the number of configurations.

Figure 4 is the result of using this approach on the event in figure 3. The left hand panels show two of the estimated trajectory configurations. From these panels it is clear that the difference between the two configurations is just that which path is followed by the electron and which by the positron has been reversed – in fact, it is typically impossible to distinguish between the electron and the positron[1]. These two configurations account for 95% of the probability of all configurations.

[1]They can be distinguished only if the positron undergoes annihilation.

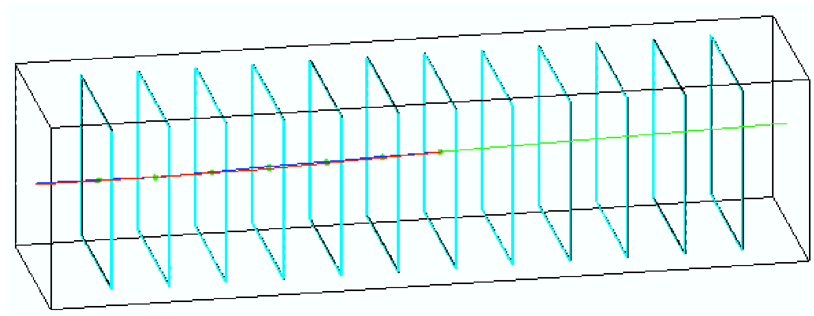

Figure 3: A gamma-ray incident on the simplified detector

The second panel shows the energy estimate for the photon. The red line is the actual energy of the simulated photon, and is well within the body of the estimated energy distribution. (Note that the histograms show the probability of *this single* photon having a particular energy.) The third panels show the estimated direction (azimuth and elevation) of the photon, and the final (rightmost) panels show the joint probability distribution over the energy of the electron and the energy of the positron. They show the estimate of how the energy of the photon is split between the two charged particles.

4 Determination of the Elementary Probabilities

It is the elementary probabilities that drive the entire event analysis system. By elementary probabilities we mean the probabilities for the physical outcomes when a given type of particle with a given energy is incident on one of the LAT conversion foils. For example, when an electron is incident on a foil, with a certain probability its direction is changed and it looses energy; with a smaller probability the above happens, and also a secondary photon is produced; with another probability a secondary electron is produced; with a smaller probability, multiple secondaries are produced.

By using a simulator similar to that in figure 3, but with only one foil, these elementary probabilities can be estimated quickly from a large number of simulations. Figure 5 shows, for the case of a muon incident on the foil, the probability of zero secondaries (top), and 1, 2, or 3 secondaries (of any type) (bottom) being produced, as a function of the energy of the incident muon. Figure 6 shows the mean energy of the secondaries produced by the incident muons, separated into the two cases of a secondary photon (bottom) and a secondary electron (top). Thus less than 1% of muon interactions produce a secondary, a proportion of which are electrons of approximately 1.5MeV. It can thus be seen that for muons it is a good approximation in this energy range to ignore all secondary particles. This will not be the case for electrons, when secondary production will be much more significant.

5 Integration into GLEAM

The software framework for developing analysis and science tools for the GLAST mission is called GLEAM, the GLAST Event Analysis Machine, and is implemented within the Gaudi software architecture developed at CERN for use in high-energy physics experiments. The components of

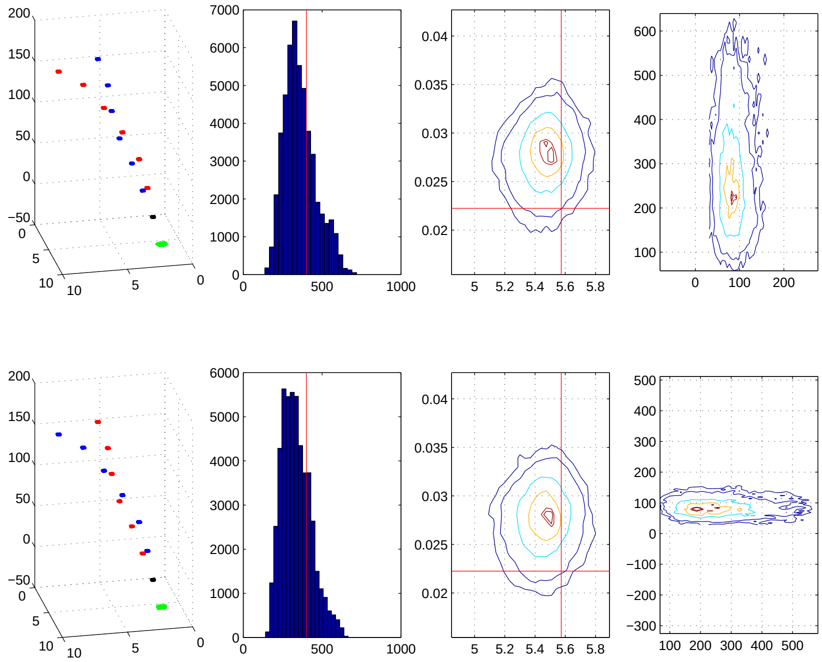

Figure 4: Analysis of the gamma-ray incident on the simplified detector. From left to right: the trajectory of the electron and positron; the estimated energy of the photon; the estimated azimuth and elevation of the photon (the psf); the energy split between the electron and positron.

the Gaudi framework we are developing are `algorithms` and `tools`. An algorithm is run by the framework once per event, and (in our case) performs the analysis. Tools are (typically) smaller pieces of code that are called many times from within an algorithm or other code. We have implemented a number of tools within the GLEAM framework.

- `IMuonScattering` - defines the interface to a tool, concrete instances of which compute the elementary probabilities for muon interactions, that is, the probability of a particular scattering angle and the probability of a particular energy loss.
- `ITrackProbability` - defines the interface to a tool which returns the probability of an entire configuration, based on scattering tools, and the configuration stored in a `PrimaryTrack` object.
- `ILikelihoodTool` - defines the interface to a tool that computes the likelihood of a `PrimaryTrack` object, based on the microstrips that fired in the event.

By defining first the *interfaces*, we can develop code in easily replaceable modules, thus facilitating the incremental development process discussed in section 2. We have implemented concrete instances of the above interface descriptions, together with other classes within GLEAM, to enable an `algorithm` which estimates the energy and direction of a muon incident on the LAT. This algorithm has been named `CalSampler`, as it uses the LAT in the manner of a sampling calorimeter.

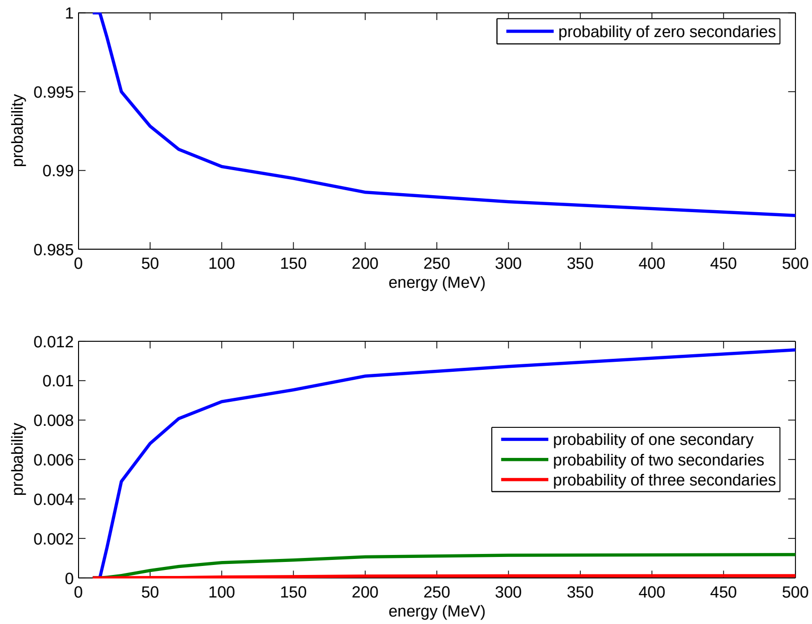

Figure 5: The probability of secondary production for muons incident on a LAT foil.

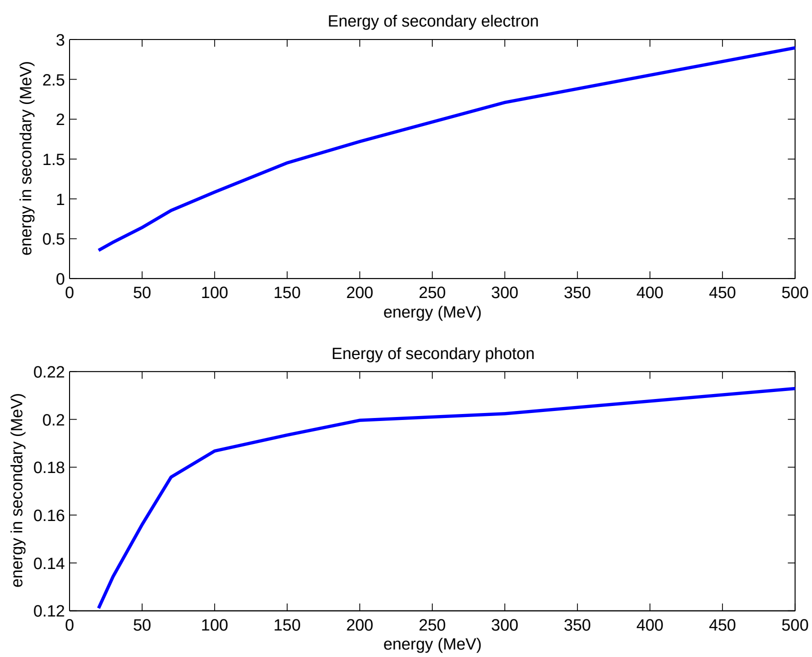

Figure 6: The energies of the secondaries produced for muons incident on a LAT foil.

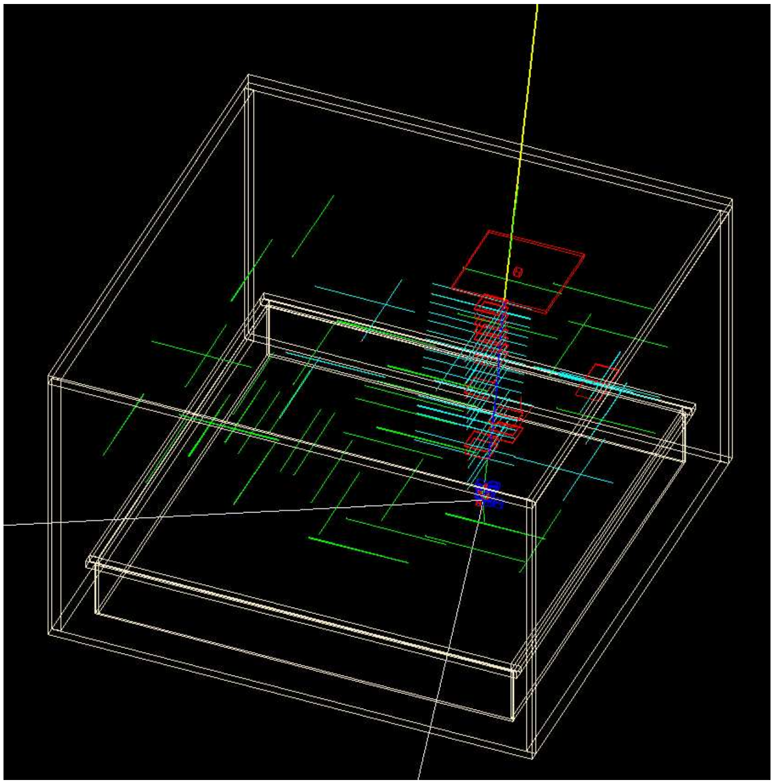

Figure 7: Simulation within GLEAM of a 100MeV muon incident on the LAT

6 Results on Simulated Muon Interactions

The GLEAM framework provides a methodology for simulating different types of particle and photons incident on the detector at different angles and energies. Figure 7 shows the simulator output when a 100MeV muon is incident on the LAT. The green lines are silicon microstrips which we have classified as noise due to their not being corresponding x- and y- measuring strips. The cyan lines represent the microstrips that are used in the estimation.

The state-space of the Markov chain that is used within the `CalSampler` algorithm consists of the energy of the incident muon; the crossing points of each of the tungsten foils; the energy loss within each foil traversed; the position at which the muon exits the detector. MCMC is run on this state space, and the resulting samples define the distribution over the energy of the particle, its trajectory, and the energy loss at each layer. Figure 8 (left) shows the energy estimated for 200 events, where each event is a 100MeV muon, incident on the LAT at an angle uniformly distributed in a 45 degree cone, originating from a random position. Only 196 events are shown, as the remaining 4 events did not trigger the detector. (The detector is "triggered" when 3 consecutive layers of silicon microstrips fire.) It can be seen that the energy estimate is nicely distributed around 100MeV (the mean value is 99MeV). The right hand graph in the figure is the energy estimates from the conventional analysis within GLEAM. The conventional analysis returned a result for only 115 of the 200 events. The energy estimates for these events have a mean of 133MeV. The methodology we have developed is thus more accurate and more widely applicable than the conventional methodology. Whilst being an interesting and useful result in itself, this result also reaffirms our proposal that extending this methodology to analyse gamma-ray events will lead to more accurate estimates from a wider range

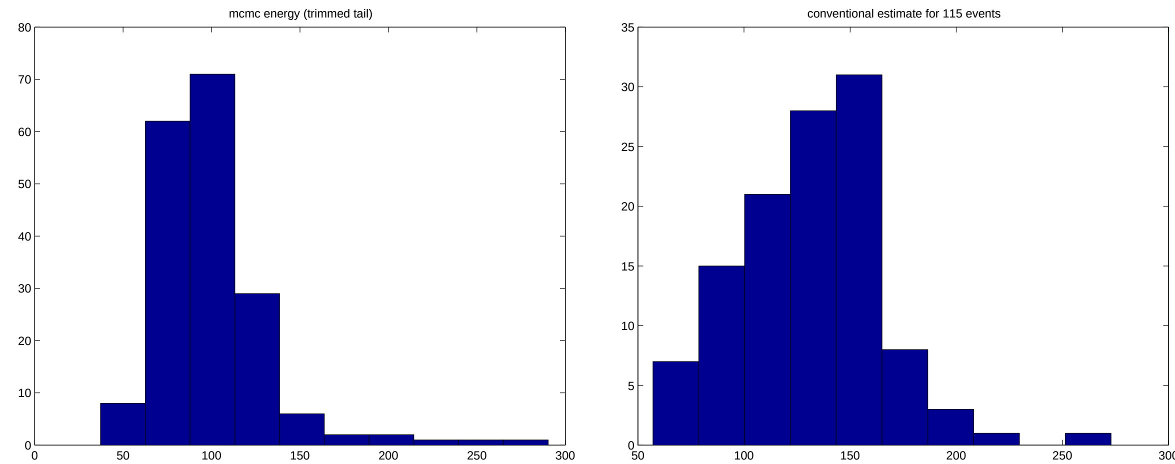

Figure 8: The distributions of energy estimates for 200 muons of 100MeV incident on the LAT. New methodology (left); conventional methodology (right)

of events. This is due to the dependence of the direction estimate on the energy estimates of the electron and proton, as described above.

7 Plans for Years 2 and 3

We present a modified work plan and milestones for years 2 and 3, based on the change to an incremental development schedule discussed earlier. All code development is now done directly within the GLEAM framework.

1. Electron/Positron Physics (3 months):

   (a) Determine via simulation the elementary probabilities for electrons incident on the LAT foils.

   (b) Implement a tool that evaluates these probabilities.

   Milestone- March 2006: Implementation of functions to evaluate the physics distributions of electrons and positrons.

2. Improved likelihood (2 months):

   (a) Study the physics of the microstrip detectors. Determine the relationships between the angle and energy of a charged particle traversing the microstrips and the number of strips that fire together with the time-over-threshold.

3. Determination of the Configurations for Incident Electrons (3 months)

   (a) Develop a tool to enumerate the possible configurations of secondary particles that could have resulted in the observed microstrip firings.

4. Analysis of Specified Electron Configurations (2 months)

   (a) Development of an MCMC sampler that explores the distribution (of energy, origin, energy loss at each layer, etc) for a specified configuration.

5. Determination of the Configurations' Probabilities (2 months)

   (a) Development of a tool that uses the output of the MCMC sampling and evaluates the relative probability of each of a set of configurations.

6. Comparative Testing (1 month)

   (a) Test the electron estimation algorithm on simulated electrons, and compare with the conventional estimation.
   (b) If possible, test on beam-test data from the actual instrument.

   Milestone- February 2007: A paper describing the methodology and results for the analysis of electrons incident on the LAT.

7. Determination of the Configurations for Incident Photons (3 months)

   (a) Expand the functionality of the previous tool to also enumerate the possible configurations for an incident photon.

8. Analysis of Specified Gamma-ray Configurations (2 months)

   (a) Expand the functionality of the previous tool to also include the two primary charged tracks and associated secondaries.

9. Determination of the Configurations' Probabilities (2 months)

   (a) Expand the functionality of the previous tool to also evaluate the relative probabilities for configurations corresponding to photons incident on the detector.

   Milestone- September 2007: Completion of the code for event analysis.

10. Complete the documentation (2 months)

    (a) Complete the documentation of the system.
    (b) Prepare journal papers (for both statistics and physics journals) describing the analysis system.

11. Analysis of events from the actual instrument (2 months)

    (a) Attempt to obtain event data from the actual instrument.
    (b) Analyze this data using the new analysis framework. Resolve any problems that become apparent.
    (c) Demonstrate the framework to interested parties in the GLAST collaboration.
    (d) Write final report.

    Milestone - January 2008: Submission of journal papers and final report.

8 Dissemination of Results

8.1 Presentations and Publications

- Seminar, INRIA, Sophia Antipolis, France, April 22nd, 2005. (INRIA is the French national computer science research institute.)
- Presentation at the meeting of the CHASC (California-Harvard AstroStatistics Collaboration) at SLAC, September 7th, 2005
- "The GLAST LAT as a Sampling Calorimeter for Muons", in preparation. To be submitted to Nuclear Instruments and Methods in Physics Research A.

8.2 Conference Committees

- Member of the organizing committee of MaxEnt 2005, the 25th Workshop on Maximum Entropy and Bayesian Methods in Science and Engineering. Session chair for the Astronomy and Physics session. Co-editor of the proceedings volume (published by AIP).
- Member of the Technical Program Committee for EUSIPCO 2006 (14th European Signal Processing Conference.

8.3 Future Invited Presentations

- "Modern Statistical Methods for GLAST Event Analysis", invited talk at "Interface 2006", 38th Symposium on the interface of statistics, computing science and its applications. May 24th-27th, 2006. Session organized by David van Dyk.